

- 1 -

PACKET DATA TRAFFIC CONTROL FOR CELLULAR
WIRELESS NETWORKS

CROSS-REFERENCE TO RELATED APPLICATIONS

This is the first application filed for the present
5 invention.

MICROFICHE APPENDIX

Not applicable.

TECHNICAL FIELD

The present invention relates to wireless networks,
. 10 and in particular to a method of increasing throughput of
packetized data traffic over a wireless network.

BACKGROUND OF THE INVENTION

In recent years there has been an exponential
increase in demand for wireless communication devices,
15 which was initiated by the development of cellular
telephones. Many new applications have been developed that
use wireless devices for transmitting digital data. These
applications include communications services such as
electronic mail and text paging, wireless World Wide Web
20 access, and wireless file transfer, etc. The use of
digitally modulated radio signals, e.g. code division
multiple access (CDMA), for packetized data transfer has
generated a demand for faster transfer rates.

Digitally modulated radio signals are subject to
25 degraded signal quality as a result of: multipath fading
caused by reflections from environmental structures such as

- 2 -

hilly terrain and/or buildings; doppler effects resulting from movement of mobile communications devices; channel interference; and background noise corruption. Multipath fading and doppler effects cause inter-symbol interference 5 (ISI) which can undesirably affect performance during data transfer. Channel interference in which signals transmitted on one channel (carrier frequency) are detected as noise on other channels, results in similar degradation of data transfer performance. In extreme cases, a 10 bit-error rate (BER) of a digital communications link can reach levels that are outside the permitted tolerance for acceptable or useful data communications.

Various methods have been proposed for maximizing the performance of wireless communications networks. For 15 example, it is known to reduce channel interference by adjusting the transmission power of base stations and wireless terminals of the network. Conventional power control techniques typically adjust wireless terminal transmission powers based on a determination of the mean or 20 average of the interference level detected at a base station. Examples of such power control techniques are provided by J. Zander, in "Distributed Cochannel Interference Control in Cellular Radio Systems," IEEE Trans. Vehic. Tech., v. 41(3), pp. 305-311 (Aug. 1992); by 25 G.J. Foschini et al., in "A Simple Distributed Autonomous Power Control Algorithm and Its Convergence," IEEE Trans. Vehic. Tech., v. 42(4), pp. 641-646 (Nov. 1993); by A.J. Viterbi, in "CDMA-Principles of Spread Spectrum Communication," ch. 4.7, pp. 113-119 (Addison-Wesley Pub. 30 Co. 1995); by S.V. Hardy, in "An Algorithm for Combined Cell-Site Selection and Power Control to Maximize Cellular

Spread Spectrum Capacity," IEEE J. Selected Areas in Comm., v. 13, no. 7, pp. 1332-1340 (Sept. 1995); and, by R.D. Yates, in "A Framework for Uplink Power Control in Cellular Radio Systems," IEEE J. Selected Areas in Comm., v.13, 5 no. 7, pp. 1341-1347 (Sept. 1995). However, the use of the simple mean of the received interference in determining the transmission power often produces a power level that is higher than necessary, causing higher than necessary levels of interference at the base stations. As a result, 10 connection capacity in such systems are still overly limited. United States Patent No. 5,732,328 (which issued to Mitra et al. on March 24, 1998) teaches an alternative method, in which transmission power is controlled on a basis of an information class (i.e. voice, audio or data) 15 of traffic on each link. The method taught by Mitra et al. permits further reductions in transmission power, with corresponding reductions in co-channel interference.

The above examples illustrate ongoing efforts to increase wireless data throughput by reducing channel 20 interference. Each of these systems seek to minimize transmission power on every link, while maintaining acceptable performance in terms of session quality of service requirements on the link in question. Thus if any one or more of carrier-to-interference (C/I) ratio; the 25 signal-to-noise (S/N) ratio; or bit error rate (BER) of any particular link falls outside of an acceptable range, then transmission power on that link is increased. However, that power increase may cause performance reductions on other links due to increased channel interference, while 30 only being effective to recover marginal performance on the link in question. Consequently, conventional methods of

Information Specified

- 4 -

reducing channel interference may have an effect of sacrificing data throughput on all links (and thus over-all data through-put) in an effort to maintain marginal data throughput on a small number of poorly-performing links.

5 Accordingly, a technique for controlling data traffic over a plurality of communications links of a wireless communications network which avoids unnecessarily diminishing overall network data throughput is highly desirable.

10 **SUMMARY OF THE INVENTION**

An object of the present invention is to provide a method of controlling data traffic in a wireless data communications network in which overall data throughput is increased by at least temporarily interrupting data transfer over poorly performing links of the network.

Accordingly, an aspect of the present invention provides a method of controlling data traffic in a wireless data communications system comprising a plurality of data communications wireless links, the method comprising the 20 steps of: identifying a poorly performing wireless link; and at least temporarily interrupting data transmission over the poorly performing wireless link. By this means, interference caused by the poorly performing wireless link is eliminated, and overall data throughput on the remaining 25 links of the wireless data communications system is improved.

Another aspect of the present invention provides a wireless data communications network comprising a base station capable of wireless communication with each of one

- 5 -

or more wireless terminals over respective wireless links,
the network comprising: computing means for identifying a
poorly performing wireless link; and control means for at
least temporarily interrupting data transmission over the
5 poorly performing wireless link.

A further aspect of the present invention provides
a base station of a wireless data communications network,
the base station being adapted for bi-directional
communications with each of one or more wireless terminals
10 over respective wireless links, the base station
comprising: computing means for identifying a poorly
performing wireless link; and control means for at least
temporarily interrupting data transmission over the poorly
performing wireless link.

15 The computing means preferably comprises means for
monitoring one or more performance parameters related to
each wireless link, and comparing each monitored
performance parameter to a respective predetermined
tolerance range. The one or more performance parameters
20 concerning each wireless link are preferably any one or
more of a quality-of-service (QOS) metric, and interference
on the link.

In an embodiment of the invention, at least one of
the performance parameters related to each wireless link
25 may be based on interference on the wireless link, and
comprises any one or more of: a signal-to-noise (S/N)
ratio; a carrier-to-interference (C/I) ratio; a
bit-error-rate (BER) user data through-put rate; and a
number of suspended frames. At least one of the
30 performance parameters related to each wireless link may be

- 6 -

an average, taken over a number n of successive bursts, of any one or more of the S/N ratio; the C/I ratio; the user data through-put; and the BER. The control means is preferably adapted to suspend transmission of a data frame 5 over the poorly performing wireless link. Transmission of the data frame may be resumed after a delay period, which may have a random length.

The computing means may also be adapted to maintain a count of suspended frames. In that case, the control 10 means is preferably adapted to suspend a communications session over the link if the number of suspended frames exceeds a predetermined threshold. The session may be restarted after a predetermined delay, which may have a random length.

15 In accordance with an embodiment of the invention, at least one of the performance parameters respecting each link is a QOS performance parameter comprising any one or more of a data transmission delay, and a number of dropped frames. The control means is preferably adapted to drop a 20 data frame transmitted over the poorly performing wireless link. The dropped frame may be retransmitted after a delay period, which may be a random delay.

The computing means may be adapted to maintain a count of dropped frames. In that case, the control means 25 is preferably adapted to drop the session if the count of dropped frames exceeds a predetermined threshold.

Preferably, the computing means is adapted to predict whether a QOS performance parameter will fail to satisfy a QOS requirement for a communications session on

- 7 -

the link. Similarly, the control means is adapted to preemptively drop a data frame being transmitted over the poorly performing wireless link.

An advantage of the present invention is that it
5 can be implemented without making any changes to existing
network infrastructure. Thus, the present invention
provides a means for improving overall data throughput and
quality of service at minimum cost to the network service
provider.

10 **BRIEF DESCRIPTION OF THE DRAWINGS**

Further features and advantages of the present invention will become apparent from the following detailed description, taken in combination with the appended drawings, in which:

15 Fig. 1. is a schematic diagram illustrating an exemplary wireless communications network usable in conjunction with the present invention; and

Fig. 2 is a flow chart illustrating exemplary steps in a process in accordance with the present invention.

20 It will be noted that throughout the appended drawings, like features are identified by like reference numerals.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

An exemplary wireless communication network 2 is
25 shown in Fig. 1. The wireless network 2 includes a plurality of base stations 4a-4d linked to a fixed-link network (not shown) such as, for example, the Public

- 8 -

Switched Telephone Network (PSTN), a broadband network employing asynchronous transfer mode (ATM) packet switching, or the Internet. Alternatively, the base stations 4 may be connected to such fixed-link networks 5 through one or more intermediate bridging components such as a switching center (not shown).

The base stations 4 provide wireless communications services to wireless terminals 6a-d located within respective geographic service areas 8a-d (i.e. cells) 10 proximate each respective base station 4a-d. Bi-directional data traffic between a base station 4 and a wireless terminal 6 within its respective cell 8 is handled by respective up-link and down-link paths of wireless data communications link 10. The size of each cell 8 may vary 15 in accordance with the respective transmission power of each base station 4 and the environment in which each base station 4 is operating. Overlap regions (not shown) typically exist between adjacent cells 8 to enable hand-off between respective base stations 4. This ensures 20 substantially uninterrupted communications of a mobile wireless terminal 6 as it travels across cell boundaries. Suitable overlap regions conform to those used in conventional wireless networks. The overlap regions are not shown for ease of illustration.

25 The base stations 4 are well known in the art and are commercially available from, for example, Nortel Networks. Similarly, the wireless terminals 6 can be, for example, mobile units, such as conventional cellular or PCS telephones, portable wireless modems for laptop computers, 30 personal digital assistants (PDA's), or stationary units, such as wireless modems associated with desktop computers

or multimedia devices. The wireless network 2 may also be a municipal area network (MAN) covering a large urban area, a small indoor network with a relatively small coverage area, such as an office building or a shopping mall, or a 5 small outdoor network such as a University campus. The method of wireless communication employed by the network 2 is, for example, a spread spectrum digital communications technique including code division multiple access (CDMA), such as that adhering to the Telecommunication Industry 10 Association Interim Standard 95. However, it is alternatively possible to use a time division multiple access (TDMA) scheme such as that adhering to the Telecommunication Industry Association Interim Standard 136, as well as other digital multiple access 15 techniques.

The network 2 supports both delay tolerant and delay intolerant wireless communication sessions between base stations 4 and the wireless terminals 6. Exemplary delay intolerant sessions include, for example, voice, 20 audio and video. Delay tolerant sessions including data transfers such as text or numerical information. Each session type typically has an associated service profile of quality of service (QoS) requirements, which may, for example, be expressed in terms of data transmission (i.e. 25 packet or frame) delay tolerance by an end-user of a wireless terminal 6.

Data transmitted during an active communications session (i.e. between a base station 4 and wireless terminal 6 over a respective wireless link 10) is in 30 Internet Protocol (IP) packets encapsulated within one or more logical link control (LLC) frames. Other data

- 10 -

protocols and transmission control formats may also be utilized. However, the present invention is particularly suitable for use in conjunction with wireless networks based on packet-switched transmission protocols.

5 In accordance with the present invention, data transmitted over poorly performing wireless links is intentionally dropped or suspended in order to improve the overall throughput of the wireless (packet-switched) network. A packet transmission control algorithm is
10 provided to identify and drop (or suspend) wireless links based on, for example, two preferred criteria: a quality of service (QOS) requirement violation, and an interference level on the link.

In the case of a QOS requirement violation, if a
15 transmitted packet or frame has already violated the applicable QOS requirement, then it can be dropped. Alternatively, if it is predicted that the frame (or packet) is very likely to violate the QOS requirement, it can be preemptively dropped.

20 In the case of an interference level on the link, where the QOS is not a factor, the transmission of a frame or packet can be suspended if the current wireless link is determined to be of such poor quality that the predicted data throughput is below a predetermined threshold.

25 The intentional interruption of data communication on poorly performing links frees band-width for use by other links, and also reduces the interference in the system. Both of these effects result in increased overall data throughput of the network by increasing the
30 performance of the remaining links, and these effects are cumulative.

- 11 -

The concept of dropping a call in a circuit-switched network is very different from that of dropping or suspending a packet or frame transmission in a packet-switched network. In the circuit-switched network, 5 dropping a call implies that the link connection between the originating and terminating users is severed, whereas this is not necessarily true in the case of a dropped frame or packet in a packet-switched network. The dropping of a call in cellular wireless circuit-switched networks usually 10 occurs in the radio link between the serving base station and the mobile user due to a harsh radio frequency (RF) environment. When designing such a system, one of the principal objectives is to minimize the number of dropped calls, which significantly affect the grade of service 15 (GOS) offered to the users by the network provider. In wireless packet-switched network systems, however, the dropping or suspension of a radio link transmission does not automatically result in a breakdown of the connection between a server and a user. The parties to the 20 communications session can still be connected via known logical set up links.

Fig. 2 illustrates steps in an exemplary control algorithm for identifying and dropping and/or suspending data transmission over poorly performing data 25 communications links in accordance with the present invention. The control algorithm is preferably instantiated during initialization of a base station 4, and runs concurrently with other communications processes. The process illustrated in Fig. 2 operates as follows.

30 At an initial step (step S1), a data communications link between the base station and a wireless terminal is

- 12 -

selected for evaluation, and performance statistics related to the selected link are obtained (step S2). The performance statistics are preferably selected from among statistics normally collected by system management processes of the data communications network 2. These may include QOS-related statistics, such as for example a total frame delay time, as well as link quality statistics such as for example the BER, the S/N ratio, the user data throughput rate, and/or the C/I ratio.

10 The obtained statistics for the link are then compared with predetermined threshold values defining minimum acceptable link performance criteria (step S3).

15 At step S4 it is determined whether the frame being transmitted over the link has failed, or will fail, to meet the minimum QOS requirements. Various algorithms may be used to implement this determination. Two possible algorithms, where the QOS-related performance statistics are related to data transmission delay are described below.

20 A first exemplary QOS algorithm assumes that an IP packet is segmented and encapsulated in N LLC frames. The maximum time permitted to transmit each LLC frame of an IP packet is calculated and the respective transmit times are stored. During transmission of each LLC frame, if the LLC frame transmit timer surpasses the maximum transmit time, 25 transmission of the IP packet is dropped (S5) and a counter (D) of IP packet drops for the selected link is incremented (S6).

Based on QoS requirements of the session, the maximum allowable delay for the transmission of an IP

packet is $t_{ip,max}$, and the maximum data throughput for the user on the given link is S_{max} bits/sec. Therefore, the minimum time (delay) for the transmission of the j^{th} LLC frame is

$$5 \quad \min(t_{\text{llc}}) = \frac{\# \text{bits in jth LLC frame}}{S_{\max}} \quad (\text{EQ } 1)$$

With this value known, the maximum allowable delay for the transmission of n^{th} LLC frame can be determined from

$$\max(t_{llc,n}) = \left[t_{ip,\max} - \sum_{i=1}^{n-1} t_{llc,i} - \sum_{j=n+1}^N \min(t_{llc,j}) \right] \quad (\text{EQ } 2)$$

If during the actual transmission of the n th LLC frame, the delay has already exceeded the above threshold (given by EQ 2), then the current LLC frame, and the remaining LLC frames of the IP packet can be dropped. This pre-emptive dropping of the IP packet is permissible because if the maximum allowable delay of the n^{th} frame is violated, then the maximum delay tolerance of the IP packet is also certain to be violated, and there is no need to continue transmitting the remaining bits of the IP packet.

A second exemplary QOS algorithm assumes that an IP packet is encapsulated in a single LLC frame, which is generally the case. The LLC frame includes a plurality of RLC blocks, as is well known in the art. In this instance, the transmission of the LLC frame can be dropped based on an average throughput of the last m RLC blocks of the LLC frame. The rate at which the LLC frame is transmitted will vary depending on the C/I ratio.

- 14 -

If it takes t_m seconds to transmit m RLC blocks, then (based on the transmissions of the most recent m RLC blocks) the average payload user data throughput can be determined from

$$5 \quad S_{avg} = \frac{\# \text{ payload bits transmitted by the } m \text{ RLC blocks}}{t_m} \quad (\text{Eq. 3})$$

It should be noted that a delay t_m includes the scheduling and decoding delays in addition to the transmission and retransmission delays. Given the average payload throughput, the time required to transmit the remaining 10 payload bits of an LLC frame can be estimated.

To do so, the number of payload bits and the maximum delay tolerance for the LLC frame are denoted by K and $t_{llc,max}$ respectively. Assuming that i payload bits have been transmitted during t_i , the estimated time or delay to 15 transmit the rest of the LLC frame, i.e. $(K - i)$ bits, is

$$\tilde{t}_{left} = \frac{\text{Remaining payload bits to be transmitted}}{S_{avg}} = \frac{(K - i)}{S_{avg}} \quad (\text{Eq. 4})$$

The maximum allowable time (delay) remaining to transmit the LLC frame is given by

$$t_{left,max} = t_{llc,max} - t_i \quad (\text{Eq. 5})$$

20 Thus, if the current allowable delay $t_{left,max}$ is less than the predicted transmit delay \tilde{t}_{left} , that is,

$$t_{left,max} < \tilde{t}_{left} \quad (\text{Eq. 6})$$

then the LLC frame is preemptively dropped, because the frame is certain to fail the QoS requirements of the 25 session. This check can be performed regularly, and as

- 15 -

often as once after the transmission of each block, or after the transmission of every m RLC blocks.

It should be noted that there are many ways to estimate the average throughput S_{avg} used to predict the 5 transmission time of the LLC frame. The method represented by EQ 3 is just one example. Other methods can be based on a different function such as a weighted average of the past m RLC blocks, or as a function of a current C/I ratio and an expected number of active users in the system.

10 Returning now to Fig. 2, if it is determined at step S4 that the current frame has either failed (or is certain to fail) the applicable QOS requirements, then the current frame is dropped (step S5). In the case of a frame being transmitted by a wireless terminal 6 over an up-link 15 path, dropping a frame can be accomplished by flushing data of the frame (or portion thereof) from buffers of the base station, and sending a control message to the corresponding wireless terminal 6 to terminate transmission of data of the dropped frame. In the case of a frame being transmitted 20 to a wireless terminal 6 over a down-link path, dropping a frame can be accomplished by sending a control message to the wireless terminal 6 indicating that the frame has been dropped.

A dropped frame counter is then incremented (step 25 S6) in order to track the number of dropped frames. The value stored in the dropped frame counter can then be compared to a threshold value (D_{max}) at step S7, to determine whether the number of frames dropped during the session (or during some other predetermined period of time) 30 has exceeded an acceptable limit. If the result of the

- 16 -

determination at step S7 is "yes", then the communications session over the selected link is dropped at step S8. Dropping the communications session may involve severing communications between parties to the communications 5 session. After the session has been dropped at step S8, a new link is selected at step S9 and processing continues as described above from step S2 with respect to the new link. Alternatively, if it is determined at step S7 that the number of dropped frames has not exceeded the maximum 10 threshold value (D_{max}), the system retries transmission of the dropped frame after a selected delay period (steps S10 and S11). Retransmission of the frame can be accomplished by sending a control signal to the wireless terminal 6 to begin sending (or receiving) the previously dropped frame, 15 starting again from the beginning of the frame. The delay period at step S10 may be a random delay. Once retransmission of the previously dropped frame has begun, the system returns to step S2 to obtain and process performance statistics concerning the frame and the link.

20 If at step S4 it is determined that the frame will not fail the applicable QOS requirements, the system proceeds to step S12 where it is determined whether excessive interference is being detected on the link. This determination is conveniently based upon any one or more of 25 the BER, the S/N ratio, the user data throughput rate, and the C/I ratio. If the link interference is determined not to be excessive at step S12, the system returns to step S9 to select and analyze a new link. If excessive interference is found at step S12, however, the frame is 30 suspended (at step S13). Suspension of the frame can be accomplished by sending a control signal to the wireless

- 17 -

terminal 6 to indicate that the sending of bits of the frame is to be stopped, but already received bits are retained within memory.

A frame suspended counter is then incremented (at 5 step S14) to count the number of frame suspensions within a predetermined time period, such as, for example, the duration of a communications session. At step S15 it is determined whether the number of frame suspensions exceeds a predetermined threshold value (FS_{max}). If the threshold 10 has not been exceeded, the suspended frame is restarted after a predetermined delay (steps S16 and S17). Restarting the frame can be accomplished by sending a control signal to the wireless station 6 to indicate that the sending of bits of the frame to the base station 4 15 should resume, starting from a point at which the frame was suspended. The delay period at step S16 can be a random delay, and may be the same or different from the delay period at step S10. Following restart of the frame at step S17, the system returns to step S2 to obtain updated 20 performance statistics for the frame and the link and then continues processing of the frame as described above.

If at step S15 it is determined that the number of frame suspensions exceeds the threshold FS_{max} , then the 25 communications session is suspended (at step S18). As with suspension of a frame, suspension of a session can be accomplished by sending a control signal to the wireless terminal 6, while retaining already received data for the session in memory. The suspended session can then be restarted (step S20) by sending an appropriate control 30 signal to the wireless terminal 6 after a predetermined delay (step S19), prior to restarting the frame (step S17).

- 18 -

Once again, the delay period at step S19 may be a random delay, and may be the same or different from the delay periods at steps S10 and S16.

The embodiment(s) of the invention described above
5 is(are) intended to be exemplary only. The scope of the invention is therefore intended to be limited solely by the scope of the appended claims.